

Claims

What is claimed is:

1. A method comprising the steps of:
5 creating a document stack from at least one word in a handwritten document;
 creating a query stack from a query; and
 determining a measure between the document stack and the query stack.
- 10 2. The method of claim 1, wherein:
 the at least one word comprises a plurality of words;
 the document stack corresponds to one of the plurality of words in the handwritten document;
 the query comprises a plurality of query words and at least one operator;
15 the query stack corresponds to one of the plurality of query words; and
 the step of determining a measure further comprises the step of, for each query stack, determining a measure between the query stack and each document stack in the handwritten document.
- 20 3. The method of claim 2, wherein each document stack comprises a plurality of document scores, and wherein the method further comprises the step of optimizing each of the document scores for the document stacks.
4. The method of claim 1, wherein the measure quantifies an amount of
25 similarity between the document stack and the query stack.

5. The method of claim 1, wherein the query is handwritten, typewritten, or partially handwritten and partially typewritten.

6. The method of claim 5, wherein the query is typewritten, and wherein the step of creating a query stack comprises creating a query stack for each query word of the query, wherein each query stack comprises a corresponding word from the query and an associated high word score for this word, and wherein each query stack comprises a plurality of other words having zero word scores associated therewith.

7. The method of claim 5, wherein the query is typewritten, and wherein the step of creating a query stack comprises creating a query stack for each query word of the query, wherein each query stack comprises a corresponding word from the query and an associated high word score for this word, and wherein each query stack comprises at least one other word having a small word score associated therewith.

8. The method of claim 1, wherein the measure is selected from the group consisting of a dot product measure, an Okapi measure, a score-based keyword measure, a rank-based keyword measure, a measure using n-grams, and a measure using edit distances.

9. The method of claim 1, where each query stack and document stack comprises a plurality of scores, wherein the measure is a dot product measure defined as follows

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{\sqrt{(\vec{q} \cdot \vec{q})(\vec{d} \cdot \vec{d})}},$$

where \vec{q} is a vector comprising scores from the query stack, and wherein \vec{d}

is a vector comprising scores from the document stack.

10. The method of claim 1, wherein each stack is not constrained to words in a vocabulary, wherein each of the words in a query stack or document stack are comprised
5 of a number of n-grams, wherein probabilities are determined for each n-gram of the query stack and document stack, and wherein the probabilities of the n-grams are used in the measure.

11. The method of claim 1, wherein each of the query and document stacks
10 comprises a plurality of words, wherein the measure uses edit distances to compare words in the query stack to words in the document stack.

12. The method of claim 1, further comprising the step of determining a document score for the handwritten document by using the measure.

13. A method comprising the steps of:
15 for each of a plurality of documents, performing the following steps:
creating a document stack from at least one word in a text document;
20 creating a query stack from a query;
determining a measure between the document stack and the query stack; and
scoring the documents based on the measure, thereby creating a document score; and
25 displaying each document whose document score meets a predetermined threshold.

14. The method of claim 13, wherein the query is a handwritten query.

15. The method of claim 13, wherein the query is a typewritten query.

5

16. A method for retrieving a subset of handwritten documents from a set of handwritten documents, each of the handwritten documents having a plurality of document stacks associated therewith, the method comprising the steps of:

10 a) creating at least one query stack from a query comprising one or more words, wherein each word is handwritten or typed;

b) selecting a handwritten document from the set of handwritten documents;

c) selecting a document stack from the selected handwritten document;

15 d) determining a measure between the at least one query stack and the selected document stack;

e) performing steps (c) and (d) for at least one document stack associated with the selected handwritten document;

20 f) performing steps (b), (c), and (d) for each handwritten document of the set of handwritten documents;

g) scoring each of the handwritten documents in the set of handwritten documents by using the query and the measures, thereby creating a number of document scores; and

25 h) selecting the subset of handwritten documents for display by using the document scores.

17. The method of claim 16, wherein step (h) further comprises the step of selecting handwritten documents that are above a predetermined threshold.

18. The method of claim 17, wherein the predetermined threshold is selected
5 from the group consisting of a rank threshold and a score threshold.

19. The method of claim 16, wherein each document stack comprises a plurality of word scores, and wherein the method further comprises the step of:

i) optimizing each of the word scores for the document stacks.

20. The method of claim 16, wherein the measure quantifies similarity between the document stack and the query stack.

21. The method of claim 16, wherein at least one of the words of the query is typewritten, and wherein step (a) further comprises the step of creating a query stack for each of the at least one words of the query, wherein each query stack comprises a corresponding word from the query and an associated high word score for this word, and wherein each query stack comprises a plurality of other words having zero word scores associated therewith.

22. The method of claim 16, wherein at least one of the words of the query is typewritten, and wherein step (a) further comprises the step of creating a query stack for each of the at least one words of the query, wherein each query stack comprises a corresponding word from the query and an associated high word score for this word, and
25 wherein each query stack comprises at least one other word having a small word score associated therewith.

23. The method of claim 16, wherein the measure is selected from the group consisting of a dot product measure, an Okapi measure, a score-based keyword measure, a rank-based keyword measure, a measure using n-grams, and a measure using edit
5 distances.

24. The method of claim 16, wherein each stack is not constrained to words in a vocabulary, wherein each of the words in a query stack or document stack are comprised of a number of n-grams, wherein probabilities are determined for each n-gram
10 of the query stack and document stack, and wherein the probabilities of the n-grams are used in the measure.

25. The method of claim 16, wherein each of the query and document stacks comprises a plurality of words, wherein the measure uses edit distances to compare words
15 in the query stack to words in the document stack.

26. A method comprising the steps of:
creating a first word stack, by using a first handwriting recognizer, from at
least one word;
20 creating a second word stack, by using a second handwriting recognizer,
from the at least one word; and
comparing the first and second word stacks with a third word stack to
determine whether a handwritten document should be retrieved.

25 27. The method of claim 26, wherein:
the at least one word is at least one handwritten word from the handwritten
document;

the first word stack comprises a first document stack;
the second word stack comprises a second document stack; and
the third word stack is a query stack determined from at least one query
word.

5

28. The method of claim 26, wherein:
the at least one word is at least one word from a query;
the first word stack comprises a first query stack;
the second word stack comprises a second query stack; and
10 the third word stack is a document stack determined from at least one
handwritten word in the handwritten document.

29. The method of claim 26, further comprising the steps of:
configuring a handwriting recognizer into a first configuration to create the
15 first handwriting recognizer; and
configuring the handwriting recognizer into a second configuration to
create the second handwriting recognizer, wherein the first and second configuration are
different.

20 30. The method of claim 29, wherein the first configuration comprises a
configuration caused by selecting a constraint from the group consisting essentially of an
uppercase letter constraint, a lowercase letter constraint, a recognize digits constraint, a
language constraint, a constraint wherein characters and words are recognized only if in a
vocabulary, and a constraint wherein characters and words are hypothesized when not in a
25 vocabulary, and wherein the second configuration comprises a configuration caused by
selecting a constraint from the group consisting essentially of an uppercase letter
constraint, a lowercase letter constraint, a recognize digits constraint, a language

constraint, a constraint wherein characters and words are recognized only if in a vocabulary, and a constraint wherein characters and words are hypothesized when not in a vocabulary.

5 31. The method of claim 26, wherein the step of comparing further comprises the step of merging the first and second word stacks to create a fourth word stack that is compared with the third word stack.

10 32. The method of claim 26, wherein the first handwriting recognizer has a first configuration, wherein the second handwriting recognizer has a second configuration, and wherein the first and second configurations are different.

15 33. The method of claim 32, wherein the first configuration comprises a configuration caused by selecting a constraint from the group consisting essentially of an uppercase letter constraint, a lowercase letter constraint, a recognize digits constraint, a language constraint, a constraint wherein characters and words are recognized only if in a vocabulary, and a constraint wherein characters and words are hypothesized when not in a vocabulary, and wherein the second configuration comprises a configuration caused by selecting a constraint from the group consisting essentially of an uppercase letter
20 constraint, a lowercase letter constraint, a recognize digits constraint, a language constraint, a constraint wherein characters and words are recognized only if in a vocabulary, and a constraint wherein characters and words are hypothesized when not in a vocabulary.

25 34. A computer system comprising:
a memory that stores computer-readable code; and
a processor operatively coupled to the memory, the processor configured

to implement the computer-readable code, the computer-readable code configured to:
create a document stack from at least one word in a handwritten document;
create a query stack from a query; and
determine a measure between the document stack and the query stack.

5

35. A computer system comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to the memory, the processor configured

to implement the computer-readable code, the computer-readable code configured to:

10 create a first word stack, by using a first handwriting recognizer, from at least one word;

create a second word stack, by using a second handwriting recognizer, from the at least one word; and

15 compare the first and second word stacks with a third word stack to determine whether a handwritten document should be retrieved.

36. An article of manufacture comprising:

a computer readable medium having computer-readable code means embodied thereon, the computer-readable program code means comprising:

20 a step to create a document stack from at least one word in a handwritten document;

a step to create a query stack from a query; and

a step to determine a measure between the document stack and the query stack.

25

37. An article of manufacture comprising:

a computer readable medium having computer-readable code means embodied thereon, the computer-readable program code means comprising:

5 a step to create a first word stack, by using a first handwriting recognizer, from at least one word;

a step to create a second word stack, by using a second handwriting recognizer, from the at least one word; and

a step to compare the first and second word stacks with a third word stack to determine whether a handwritten document should be retrieved

10

206T20 "T426200T